

## GALE Machine-Translation Post-editing Guidelines (v. 2.5, 6/27/05)

The contents of this box is the guidelines for editing the machine translation output (MT output). You should memorize them. Do not spend a long time trying to find the absolute minimal edit, the very best word, or “the felicitous phrase”.

Make the MT output have the *correct meaning*, be *readily understandable*, and really be *English*.

- (1) The goal is to edit the MT output so that it has **the same meaning** as the human translation.
- (2) Edit the MT output so that it is **understandable**.
- (3) Punctuation must be understandable, and sentence-like units must have sentence-ending punctuation. But do not insert, delete, or change other punctuation merely to follow optional traditional rules about what is “strictly correct.”
- (4) If words/phrases/punctuation in the MT output or the reference human translation are completely acceptable, prefer them over substitutions.
- (5) Dates, as well as the commas and decimal points in numbers, should be formatted according to U.S. conventions (for example, convert 23-2-2004 to 2-23-2004).

A comment about fluency or stylistic felicity: One can assign grades or categories to the fluency of a translation. One possible set of categories would be the following.

- (1) Incomprehensible mess.
- (2) Ugly but minimally understandable
- (3) Non-native English. English with errors that eleven year old (sixth grader) native speakers of English do not normally make when speaking. Language spoken by young children who have not yet learned all the syntax of English. **Fundamental errors are present. Not quite English.**
- (4) Something that an eleven year old native speaker of English would be satisfied to produce when speaking to another eleven year old native speaker of English. There are no important syntax errors. Verbs are in the correct tense, and their arguments are correct. Words are used with the correct meaning. The speaker knows English, but it’s not necessarily very polished. Casual conversation between adult native speakers of English. **No fundamental errors. Can reasonably be called English.**
- (5) English that a high-school teacher who is a native speaker of English might produce when teaching, or that a typical newspaper news article might contain. Typical prepared business talks and scientific/technical talks. Talk show hosts. Reasonably educated adult native speaker of English who already knows what they intend to get across or who is writing a first draft letter to an ordinary friend.
- (6) Typical professionally edited non-fiction books intended for the general public. Typical mass-market fiction. Network news *read* by the anchor person. This category is written English that is not a first draft.
- (7) English that a highly polished speaker such as a well-spoken politician or minister might produce in a prepared speech written in advance and read from notes/script, or which might appear in a well-polished essay or a literary novel in English. English in this category has been re-written repeatedly and polished to the highest degree by someone with a well-developed sense of style.

As you edit the MT output, turn it into something that is category 4 or 5 (remove fundamental errors, so that it is really English and has no real errors). The stylistic fluency of the edited version doesn’t have to be as good as the reference translation (the reference is probably category 5 or 6).